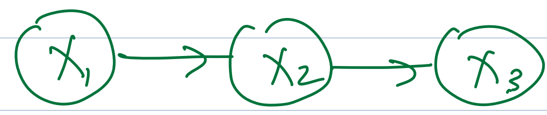
Multiple factorizations

Generative model:



True DAG.

i.e., $X_1 \perp\!\!\!\perp X_3 \mid X_2$.

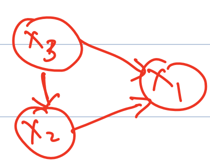
Let $p(x_1, x_2, x_3) = p(x_1)\, p(x_2 \mid x_1)\, p(x_3 \mid x_2)$

be the true underlying distribution.

Given $p(x_1, x_2, x_3)$, we can have an alternate factorization, using chain rule:

$$p(x_1, x_2, x_3) = q(x_3)\, q(x_2 \mid x_3)\, q(x_1 \mid x_3, x_2)$$

$\hookrightarrow (*)$



Alternate model using chain rule

Since we know that $X_1 \perp\!\!\!\perp X_3 \mid X_2$, we know that there exists functions s.t.

$$p(x_1, x_2, x_3) = \phi_{32}(x_3, x_2)\, \phi_{12}(x_1, x_2).$$

Comparing the factorizations, this indicates that the above alternate factorization (*) should satisfy

$$q(x_1 | x_3, x_2) = q(x_1 | x_2), \text{ i.e., the}$$

alternate graph should have a '0 weight' edge :

Thus, in this example, the true graph and the alternate graph has the same skeleton (i.e., undirected edges).

Question: In general, can we use CI relationships to uniquely determine the correct DAG (i.e, correct directions for edges) ?

No! Counter-example: Rain and wetness from Notes 1.

$$X_1 \to X_2 \quad \text{vs.} \quad X_1 \leftarrow X_2$$

$$p(x_1, x_2) = p(x_1)\, p(x_2 | x_1) \quad \text{vs.} \quad p(x_1, x_2) = q(x_2)\, q(x_1 | x_2)$$

We cannot distinguish between these two factorizations from observational data. This again motivates the study of interventions to determine structure (here, causal relationships).